

DOI:10.16356/j.1005-2615.2016.05.001

时空轨迹群体运动模式挖掘研究进展

吉根林 孙鸿艳 赵斌

(南京师范大学计算机科学与技术学院, 南京, 210023)

摘要: 群体运动模式是时空轨迹模式挖掘的重要内容, 用于发现群体运动规律、群体运动趋势以及群体事件。本文首先对群体运动模式建模和群体运动模式挖掘两个层面存在的问题与挑战进行了阐述。接着, 对群体运动模式进行了分类, 将其分为有领导者运动模式、伴随模式、突变运动模式、流行运动模式、聚集运动模式和发散运动模式。最后, 介绍了各种群体运动模式之间的区别与联系, 对各种群体运动模式挖掘算法思想进行了综述。

关键词: 群体运动模式; 时空轨迹; 群体事件; 轨迹模式挖掘

中图分类号: TP181

文献标志码: A

文章编号: 1005-2615(2016)05-0615-10

Research Progress in Group Moving Patterns Mining of Spatio-Temporal Trajectories

Ji Genlin, Sun Hongyan, Zhao Bin

(School of Computer Science and Technology, Nanjing Normal University, Nanjing, 210023, China)

Abstract: Group moving pattern is an essential problem in spatio-temporal trajectory pattern mining which is used to get group moving rules, group moving trend and group events. The problems and challenges of group moving pattern modeling and mining are analyzed. Group moving patterns are classified as having leaders moving pattern, company moving pattern, mutant moving pattern, popular moving pattern, aggregation moving pattern, and divergence moving pattern. The difference and connection between different group moving patterns and the idea of algorithms for mining group moving patterns are introduced.

Key words: group moving pattern; spatio-temporal trajectory; group event; trajectory pattern mining

时空轨迹是地理空间位置维度和时间维度上的曲线, 用于表示移动对象在一段时间内的地理空间位置变化, 每条时空轨迹由一系列时空采样点构成, 每个采样点记录了移动对象的采样时间、采样位置和方向等信息, 刻画了移动对象的历史移动行为。随着收集到的移动对象的轨迹数据量的快速

增长, 对这些轨迹数据的分析挖掘需求也逐渐增长, 其中运动模式挖掘是时空轨迹挖掘领域中一个发展迅速的前沿研究课题。通过轨迹模式挖掘算法, 可以发现轨迹大数据中有价值的模式, 从而分析移动对象的运动趋势和运动规律。

Jeung 等人^[1]将轨迹模式按粒度划分为个体

基金项目: 国家自然科学基金 (41471371) 资助项目。

收稿日期: 2016-07-01; **修订日期:** 2016-09-01

作者简介: 吉根林, 男, 教授, 博士生导师, 主要研究方向为数据挖掘技术及其应用。江苏省高校“青蓝工程”中青年学术带头人, 曾获江苏省科学技术奖、江苏省教学成果奖。累计发表论文 90 余篇。

通信作者: 吉根林, E-mail: glji@njnu.edu.cn。

引用格式: 吉根林, 孙鸿艳, 赵斌. 时空轨迹群体运动模式挖掘研究进展[J]. 南京航空航天大学学报, 2016, 48(5): 615-624. Ji Genlin, Sun Hongyan, Zhao Bin. Research progress in group moving patterns mining of spatio-temporal trajectories[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2016, 48(5): 615-624.

模式^[2]和群体模式。群体模式按照涉及的时长分为单时间片上的群体共现模式和多时间片上的群体运动模式。单时间片上的群体共现模式能够挖掘某个时间片上的群体对象运动状态,不能体现出群体对象的运动规律;多时间片上的群体运动模式能够有效反映群体移动对象的运动规律、运动趋势,也可以反映各种群体事件的发生。因此多时间片上的群体运动模式在很多领域有着广泛的应用。例如拼车应用系统^[3]通过联系有相同路线的人来降低车费,物流系统为了实现更好的规划需要检查同一组的送货车。推荐系统^[4]通过挖掘群体轨迹了解包含目标客户的群体对象的行为来为目标客户提供广告以及定制的基于位置的服务。科学家通过研究群体对象的运动(大黄蜂,蜜蜂,鱼)寻找成群的动物和动物的习惯模式。团队体育赛事^[5](足球,棒球,橄榄球)中,通过研究群体对象的运动轨迹,可以更好地了解该项体育运动,分析运动中使用的策略,提取使用策略的位置和时间。交通分析应用系统利用群体对象的移动轨迹研究人群的运动趋势和运动规律。群体轨迹表明群体对象的历史移动行为,通过识别群体对象的运动轨迹可以更好地了解群体对象,更好地管理运输基础设施。群体异常行为可用于发现异常的道路,如道路施工、路段坍塌,危险的行为,恐怖袭击等。因此,群体运动模式以及群体运动模式挖掘研究有着重要的理论意义和实际应用价值。

1 问题与挑战

群体运动模式是时空轨迹模式挖掘研究中具有挑战性的问题。群体运动模式研究的问题与挑战是群体运动模式建模和群体运动模式挖掘。

群体运动模式建模是群体运动模式挖掘的前提,只有先对群体运动模式建立适合的模型,才能根据模型挖掘海量时空轨迹数据中群体运动行为,因此,对群体运动模式建立适合的模型至关重要。为了对群体运动模式进行建模,需要先找到有意义的、能反映群体事件的、群体运动规律的群体对象的运动轨迹,然后提取群体运动轨迹特征,最后根据提取的群体运动轨迹特征建立群体运动模型。但是随着位置定位技术和通信技术的提高,海量的时空轨迹数据被采集,从海量的时空轨迹数据中找到有意义的、能反映群体事件、群体运动规律的群体对象移动轨迹具有极大的困难。另外,群体运动模型直接关系到群体运动模式挖掘,因此如何建立简单且能描述群体运动行为的群体运动模型,以降

低模式挖掘难度具有一定的挑战性。

在给出群体运动模式模型之后,需要根据给定的群体运动模式模型从海量轨迹中挖掘群体运动模式。群体运动模式模型复杂度高,涉及的群体对象多且具有不确定性,因此模式挖掘算法的提出具有一定的挑战性。由于时空轨迹数据是由定位设备不断地传输到服务器,若将每个时刻到达的时空轨迹数据都存储到历史轨迹数据库中,并从历史轨迹数据库的所有轨迹数据中挖掘群体运动模式,则计算量太大且不具有实际应用价值,因此,增量更新问题的解决非常必要,且具有一定难度。有实时性要求的应用或者有关流式数据的应用对模式挖掘算法的高效性有着更高的要求。

2 群体运动模式建模

群体运动模式建模是群体运动模式挖掘研究的前提,只有建立合适的群体运动模式,才能从海量时空轨迹移动数据中挖掘群体运动,群体运动模式按运动形态可分成多种,根据目前研究现状,具体分类如图1所示。

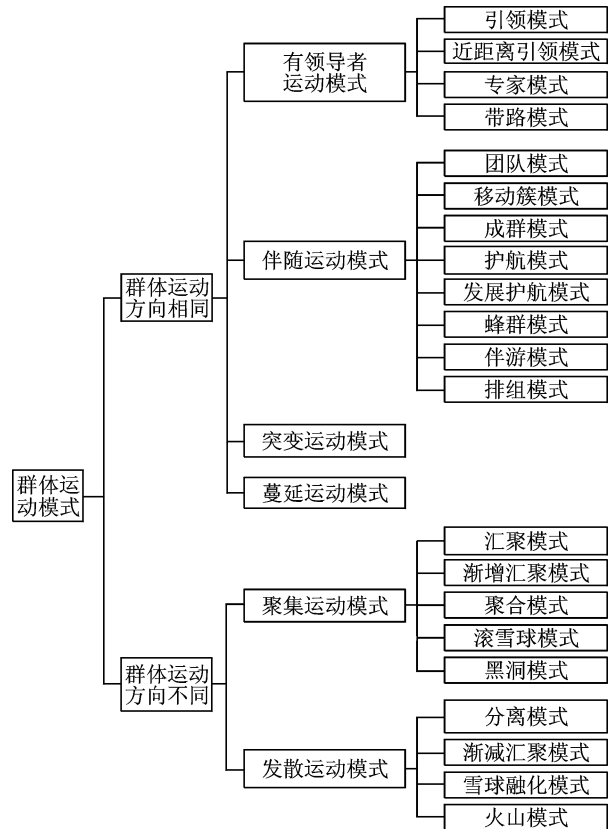


图1 群体运动模式分类

Fig. 1 Group moving pattern classification

从图1中可以看出,群体运动模式主要分为方向相同的群体运动模式和方向不同的群体运动模

式。方向相同的群体运动模式主要包括有领导者运动模式、伴随运动模式、突变运动模式和蔓延运动模式;方向不同的群体运动模式主要包括聚集运动模式和发散运动模式。有领导者运动模式是指群体中有领导者、有跟随者的运动模式,该模式的特点是跟随者与领导者有相同或相似运动属性,如速度、方向、加速度或空间位置等,但是与领导者有明显的时间滞后,可以为商业决策等提供支持;伴随运动模式是指群体对象彼此空间位置接近地移动一段时间的运动模式;突变运动模式是运动属性突变的运动模式,如方向突然改变、运动速度突然改变或者加速度突然改变等,该模式的挖掘用于发现异常路段等;蔓延运动模式是指越来越多的移动对象表现出相同的运动属性,如方向、速度等;聚集运动模式是群体对象向某区域聚集的运动模式,该模式的挖掘可以用于提前发现高密度区域,避免踩踏等有害事件的发生;发散运动模式是群体对象从某个区域向区域外移动的运动模式,可以用于挖掘恐怖袭击等异常事件。上述 6 种群体运动模式并不是完全无关,而是存在部分相关,例如伴随运动模式中群体对象突然改变运动属性时,伴随运动模式也可以是突变运动模式。

2.1 有领导者运动模式

有领导者运动模式又分为 4 种模式,包括引领模式 (trendsetter)、近距离引领模式 (greese flock)、专家模式 (guru) 以及带路模式 (leadership),如表 1 所示。

表 1 有领导者运动模式

Tab. 1 Having leaders moving pattern

模式	含义	不同点	提出时间/年
引领模式	多个对象表现某种运动属性,如速度、方向或加速度等,之后有群体对象表现出该运动属性	没有空间位置的限制	2002
近距离引领模式	近距离的引领模式	要求近距离空间	2002
专家模式	远距离的引领模式	要求远距离空间	2002
带路模式	有空间限制的引领模式	有空间限制	2005

2000 年,Imfeld 等人研究了野生动物的运动^[6],2001 年,Frank 研究了社会经济与人类运动之间的关系^[7],在此基础上,2002 年,Laube 等人提出了多种群体运动模式^[8],包括引领模式在内。引领模式定义为多个对象表现某种运动属性,如速度、方向或

加速度等,之后有群体对象表现出该运动属性。

近距离引领模式定义为近距离的引领模式,专家模式定义为远距离的引领模式,由 Laube 等人^[8]于 2002 年在引领模式的基础上提出,近距离引领模式和专家模式在引领模式的基础上限制了空间位置,不同点在于近距离引领模式用于挖掘近距离的引领模式,专家模式用于挖掘远距离的引领模式。

Laube 等人^[9]于 2005 年在引领模式的基础上提出带路模式,用于挖掘有空间限制的引领模式。

2.2 伴随运动模式

伴随运动模式又包含多种模式,包括团队模式 (Group)、移动簇模式 (Moving cluster)、成群模式 (Flock)、护航模式 (Convoy)、发展护航模式 (Evolving convoy)、蜂群模式 (Swarm)、伴游模式 (traveling companions) 以及排组模式 (Platoon),如表 2 所示。2003 年,Wang 等人^[10]首次提出团队模式,要求群体对象彼此之间的距离在一定阈值范围内且连续时间达到一定阈值,且总的的时间达到一定阈值。2013 年,Feuerhake 等人^[11]提出了频繁序列团队模式,用于识别具有重复性的防守、进攻等运动模式,2015 年,Wang 等人^[12]提出了频繁团队模式的概念,该模式要求群体对象规模达到一定阈值,且总的的时间达到一定阈值。并且针对不确定轨迹提出可能频繁团队模式的概念,即计算群体对象总是出现在一个簇中并且时间大于一定阈值的概率大于一定阈值的团队模式群体。

由于静态密集区域识别^[13],发现一起运动一段时间的群体对象^[14],对少部分对象更新有效的研究^[15]不适用于可能所有对象都更新的动物迁徙的情况,因此,Kalnis 等人^[16]于 2005 年提出移动簇模式,要求相邻两个时刻的簇中共同对象规模与所有对象规模的比值大于阈值,用于发现有对象加入或离开的群体运动模式。该模式应用广泛,如研究动物群体移动的演化,监督并行移动的部队等。

2002 年 Laube 等人^[8]提出的关联运动中,成群模式被定义为受限空间上群体对象表现出相同的运动属性,如速度、方向或加速度等。2006 年,Benkert 等人^[17]重新定义了成群模式,成群模式要求规模大于一定阈值的群体对象在一个固定半径的圆形区域内持续时间达到一定阈值。

成群^[18]模式不能发现给定半径的圆形区域外的群体对象。移动簇模式^[16]不能发现群体对象固定的群体运动。因此,Jeung 等人^[19]于 2008 年首次提出护航模式,要求密度相连的群体对象一起移

表2 伴随运动模式

Tab. 2 Company moving pattern

模式	含义	提出时间/年
团队模式	群体对象彼此之间的距离在一定阈值范围内连续时长达到一定阈值,且总的时长达到一定阈值	2003
移动簇模式	相邻两个时刻的簇中共同对象规模与所有对象规模的比值大于阈值	2005
成群模式	规模大于一定阈值的群体对象在一个固定半径的圆形区域内持续时间达到一定阈值	2006
护航模式	密度相连的群体对象一起移动至少一定阈值的连续时间戳	2008
发展护航模式	达到给定阈值的无中途加入或离开的对象在每个时间戳密度相连,零个或多个中途加入或离开的对象在一定阈值长度的时间窗口内有持续一定阈值时间的密度相连	2010
蜂群模式	大于一定阈值规模且密度相连的群体对象一起移动的时间戳达到一定阈值,且时间上不要求连续	2010
伴游模式	规模大于一定阈值且密度相连的群体对象一起移动的时间达到一定阈值	2012
排组模式	规模达到一定阈值的群体对象连续密度相连的时间达到一定阈值,且总的密度相连移动时间达到一定阈值	2015

动至少一定阈值的连续时间戳,该模式能够为汽车租赁等提供决策支持。

由于护航模式^[19]不能挖掘有对象加入或离开的群体运动,与现实情况不符,因此 Aung 等人^[20]于 2010 年提出发展护航模式,用于挖掘有对象加入或离开行为的护航模式,要求达到给定阈值的无加入或离开行为的对象在每个时间戳密度相连,零个或多个有加入或离开行为的对象在一定阈值长度的时间窗口内有持续一定阈值时间的密度相连。可用于交通规划和野生动物研究。

移动簇模式,成群模式和护航模式都要求达到一定阈值的连续时间戳,不适用于挖掘时间戳不连续的运动模式。2010 年, Li 等人^[21]提出了蜂群模式,放松了对连续时间戳的要求,要求大于一定阈值规模且密度相连的群体对象一起移动的时间戳达到一定阈值,且时间上不要求连续。可以用于动物行为的深度研究,车辆控制等。

目前对成群模式,护航模式,蜂群模式的研究,主要是基于静态数据库,并且需要多次扫描数据,或者不能以增量方式输出结果。为了在流式数据上挖掘群体伴随运动群体, Tang 等人^[22]于 2012 年提出了伴游模式,用于挖掘流式数据上的规模大于一定阈值且密度相连的群体对象一起移动的时间达到一定阈值,用于科学研究、交通管理、军事侦察。

成群模式、护航模式等严格限制时间,蜂群等模式非常放松对时间的要求,严格限制时间会导致部分模式的丢失,而非非常放松对时间的要求会挖掘噪声模式。因此为了不丢失部分运动模式,同时也不挖掘噪声模式, Li 等人^[23]于 2015 年提出排组模式,要求规模达到一定阈值的群体对象连续密度相连的时间达到一定阈值,且总的密度相连移动时间达到一定阈值。该模式可以应用于多个场景,共同

路径识别、提供对动物迁徙更加深入的理解以及协助警察识别嫌疑犯群体运动。

各种伴随运动模式的相同点是要求群体对象彼此之间地理空间位置接近。为了描述各种伴随运动模式之间的不同点,从 5 个方面进行比较,包括是否要求群体对象之间密度相连、时间上是否需要达到连续性要求、是否有一定时间阈值要求、群体对象规模是否要达到一定阈值要求、模式时间内各个时间戳的群体对象是否完全相同,如表 3 所示。

表3 伴随运动模式不同点

Tab. 3 Difference between company moving patterns

运动模式	密度相连	连续	时长	规模	对象相同
团队模式		✓	✓	✓	✓
移动簇模式	✓	✓		✓	
成群模式		✓	✓	✓	✓
护航模式	✓	✓	✓	✓	✓
发展护航模式	✓	✓	✓	✓	
蜂群模式	✓		✓	✓	✓
伴游模式	✓	✓	✓	✓	✓
排组模式	✓	✓	✓	✓	✓

团队模式不同于其他模式的是要求群体对象彼此之间的距离在一定阈值内,而不要求密度相连;成群模式与其他模式最大的不同是要求群体对象在一个半径固定的圆形区域内;移动簇模式要求相邻簇中存在超过一定阈值的共同对象;发展护航模式要求存在超过一定阈值的相同对象;伴游模式的研究对象为流式数据;排组模式放松了对时间的要求,不要求全局连续,只要求局部连续时间达到一定阈值;蜂群模式则完全放松了对时间要求。

2.3 聚集运动模式

聚集运动模式用于发现聚集的群体对象,若空间区域中存在的群体对象密度超过一定程度,则容

易发生踩踏等事件,因此对群体聚集运动模式的建模与挖掘非常重要,目前提出的各种聚集运动模式如表4所示,包括汇聚模式(Convergence)、渐增汇聚模式(Real-Gpattern \geq)、聚合模式(Gathering)、滚雪球模式(Snowball+)以及黑洞模式(Black holes)。

表4 聚集运动模式

Tab. 4 Aggregation moving pattern

模式	含义	提出时间/年
汇聚模式	向空间中某区域移动的群体对象	2002
渐增汇聚模式	群体对象密度相连时间大于一定阈值,且后一个时刻密度相连的群体对象包含前一个时间戳密度相连的群体对象,最后一个时间戳密度相连的对象规模大于第一个时间戳密度相连的对象规模	2012
聚合模式	密集、密度相连且达到一定规模的群体对象在一个稳定的区域持续时间达到一定阈值,且有一定数量的主要参与者	2013
滚雪球模式	相邻两个时间戳对应的密度相连的共同对象的个数大于一定阈值,时间大于一定阈值,且对象增加的速度大于一定的阈值	2014
黑洞模式	人群进流量与出流量之差大于一定阈值的子图	2015

2002年,Laube提出汇聚模式^[8],该模式是最早提出的聚集运动模式,用于挖掘向空间中某区域移动的群体对象,标志着聚集运动模式研究的开始。

Hai等人^[24]于2012年提出渐增汇聚模式,要求群体对象密度相连时间大于一定阈值,且后一个时刻密度相连的群体对象包含前一个时间戳密度相连的群体对象,最后一个时间戳密度相连的对象规模大于第一个时间戳密度相连的对象规模。

Zheng等人^[25]于2013年提出聚合模式,要求密集、密度相连且达到一定规模的群体对象在一个稳定的区域持续时间达到一定阈值,且有一定数量的主要参与者。用来检测大规模商业推广、集会、游行、抗议、交通堵塞等各种群体事件。

Guo等人^[26]于2014年在移动簇模式,成群模式,护航模式,伴游模式和蜂群模式的理论研究基础上,提出滚雪球模式,要求相邻两个时间戳对应的密度相连的共同对象的个数大于一定阈值,时长大于一定阈值,且对象增加的速度大于一定的阈值。用于通过移动对象的异常行为进行疾病预测等。

Li等人^[27]于2010年在图的理论基础研究之上提出了网络图上的黑洞模式和火山模式,认为黑洞模式是只有进流量没有出流量的子图,火山模式是只有出流量没有进流量的子图。2012年,Li等人^[28]提出了新的黑洞模式和火山模式,即只有出流量平均数大于进流量的平均数的子图才是黑洞模式,出流量平均数小于进流量的平均数的子图是火山模式。2015年,Hong等人^[29]在Li等人的研究基础上提出黑洞模式,该模式是人群进流量与出流量之差大于一定阈值的子图,且要求该子图表示的地理空间范围小于一定阈值,用于挖掘聚集运动。该运动模式可以用于检测疾病、灾难性事故等异常事件,保持公共安全,帮助形成更好的城市计划。

渐增汇聚模式、滚雪球模式和黑洞模式主要是根据区域范围内群体对象规模的变化判断是否是群体聚集运动,挖掘的是动态的聚集过程,聚合模式挖掘的是静态的聚集状态。渐增汇聚模式的特点是要求前一个时间戳参与聚集的群体对象在后一个时间戳依然参与聚集运动。滚雪球模式只要要求前一个时间戳参与聚集的部分对象在后一个时间戳依然参与聚集运动。黑洞模式根据图上的进流量和出流量的差值判断是否是聚集运动。

2.4 发散运动模式

发散运动模式可以用于发现发散的群体运动,如活动结束后或恐怖袭击时群体对象从某个空间向外扩散等。各种发散运动模式如表5所示,包括分离模式(Divergence)、渐减汇聚模式(Real-Gpattern \leq)、雪球融化模式(Snowball-)以及火山模式(Volcanos)。

2002年,Laube提出分离模式^[8],用于挖掘从某

表5 发散运动模式

Tab. 5 Divergence moving pattern

模式	含义	提出时间/年
分离模式	从某区域向区域外移动的群体对象	2002
渐减汇聚模式	群体对象密度相连时间大于一定阈值,且前一个时刻密度相连的群体对象包含后一个时间戳密度相连的群体对象,最后一个时间戳密度相连的对象规模小于第一个时间戳密度相连的对象规模	2012
雪球融化模式	相邻两个时间戳对应的密度相连共同对象个数大于一定阈值,时长大于一定阈值,且对象减少的速度大于一定的阈值	2014
火山模式	人群出流量与进流量之差小于一定阈值的子图	2015

区域向区域外移动的群体对象,该模式是最早提出的发散运动模式,标志着发散运动模式研究的开始。在该模式提出之后,提出了多种群体发散运动模式。

Hai 等人^[24]于 2012 年提出渐减汇聚模式,要求群体对象密度相连时间大于一定阈值,且前一个时刻密度相连的群体对象包含后一个时间戳密度相连的群体对象,最后一个时间戳密度相连的对象规模小于第一个时间戳密度相连的对象规模。

Guo 等人^[26]于 2014 年提出雪球融化模式,要求相邻两个时间戳对应的密度相连的共同对象的个数大于一定阈值,时长大于一定阈值,且对象减少的速度大于一定的阈值。

2015 年, Hong 等人^[29]在 Li 等人的研究基础上提出火山模式,用于挖掘发散运动模式,该模式是人群出流量与进流量之差小于一定阈值的子图,且要求该子图表示的地理空间范围小于一定阈值。

汇聚模式与众不同的是考虑某个时间段群体对象的运动趋势,渐减汇聚模式和雪球融化模式之间的区别在于对连续时间戳参与发散运动的共同对象规模要求不同,渐减汇聚模式前一个时间戳参与发散运动的群体对象完全包含后一个时间戳参与发散运动的群体对象,而雪球融化模式只要求部分相同。火山模式根据图的人群进出流量之差判断是否是发散运动模式。

2.5 群体运动模式研究路线

根据各种群体运动模式的发展的脉络总结它们之间的关系,如图 2 所示。

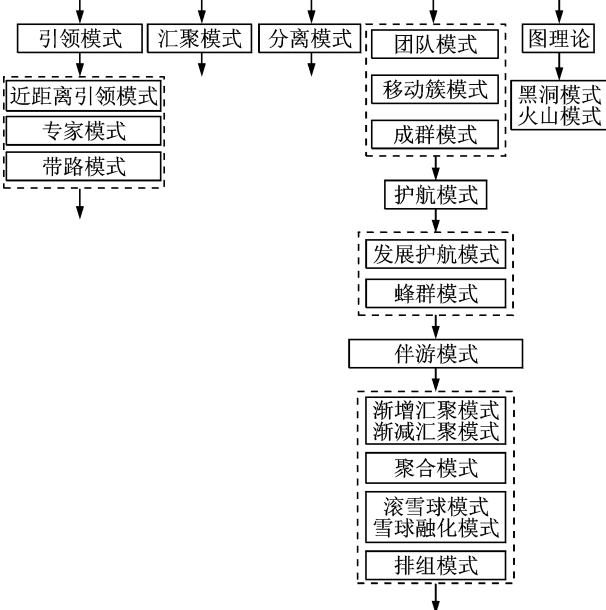


图 2 各种群体运动模式之间的关系

Fig. 2 Relationship between different group moving patterns

图 2 中箭头下方的群体运动模式是在箭头上方的群体运动模式的理论研究基础上提出,并按照时间顺序从左到右排列。

3 群体运动模式挖掘算法

3.1 有领导者运动模式挖掘

(1) 基于网页排名 (PageRank) 的模式挖掘算法

Solera^[30]等人于 2015 年提出基于贸易 (Tarde) 理论的模式挖掘方法,建立跟随者指向领导者的有向图,通过网页排名算法计算网页排名值,根据网页排名值判断是否是领导者或跟随者模式。

(2) 基于区域查询的模式挖掘算法

Andersson 等人^[31]于 2007 年提出基于区域查询的模式挖掘方法,近似范围跟随者子集查询方法,建立近似前向三角形区域,根据这些三角形区域快速查询跟随者,进而判断是否满足模式要求。

3.2 伴随运动模式挖掘

群体伴随运动模式挖掘算法主要分为 3 种,基于关联规则的模式挖掘算法、基于聚类的模式挖掘算法和基于区域查询的模式挖掘算法,如表 6 所示。

表 6 伴随运动模式挖掘算法

Tab. 6 Company moving pattern mining algorithm

类型	模式	提出时间/年	文献
基于关联规则的模式挖掘算法	团队模式	2003	[9]
		2005	[31]
		2008	[32]
基于聚类的模式挖掘算法	移动簇模式	2008	[15]
	护航模式	2008	[18]
		2008	[33]
	2009	[34]	
	发展护航模式	2012	[19]
	蜂群模式	2010	[35]
	2010	[20]	
伴游模式	2012	[21]	
基于区域查询的模式挖掘算法	成群模式	2015	[22]
		2006	[16]
		2006	[36]
		2007	[37]

(1) 基于关联规则的模式挖掘算法

2003 年, Wang 等人^[10]提出了基于 Apriori 关联规则挖掘算法的有效组模式挖掘算法 (Apriori-like algorithm for mining valid group patterns, AGP) 以及基于 FP-Growth 算法的有效组图结构挖掘算法 (Algorithm based on valid group graph

data structures, VG-Growth) 的挖掘团队模式。

上述两个算法存在两个缺点,即为了保证位置信息的精确,对移动对象位置信息的采样频率必须很高,这样就会使得数据库变得非常庞大;由于基站时钟的差异,在现实中,对用户的位置信息的采集几乎是不可能完全同步。为了克服以上缺点, Hwang 等人^[32]在 2005 年提出了用轨迹模型来表示物体的移动,即一条轨迹可以分解成一个线性函数的集合,每个线性函数的时间区间都不相交。基于 AGP 算法和 VG-growth 算法,他们提出两个改进算法:基于轨迹的组模式挖掘算法(Apriori trajectory-based group pattern mining, ATGP) 和遍历式 VG-Growth 算法(Traversal VG-growth, TVG-growth),用于挖掘团队模式。

2008 年, Wang 等人^[38]扩展了 AGP 和 VG-growth 算法,提出了基于 Apriori 的最大有效组挖掘算法(Apriori-based algorithm for mining maximal valid groups, AMG) 和生成最大有效组的 VG-Growth(an extension of VG-Growth for generating maximal valid groups, VGMax) 两个算法来挖掘最大团队模式。

(2) 基于聚类的模式挖掘算法

2008 年, Kalmis 等人^[16]提出了名为移动簇一算法(MC1), 在每个时刻对移动物体进行聚类, 对相邻时刻的簇求交集来判断它们是否包含足够数目的公共对象。但是由于对每个时刻聚类产生较大的计算量, 计算性能较低, 因此他们在 MC1 的基础上提出移动簇二算法(MC2), 在每个时刻进行聚类后, 对相邻簇求交集时增加剪枝操作来降低计算量; 为了进一步改进性能, 在 MC2 基础上提出了一个近似算法移动簇三算法(MC3), 通过减少具有噪声的基于密度的聚类算法(Density-based spatial clustering of applications with noise, DBSCAN) 聚类时物体的数目来降低运行时间。

Jeung 等人^[19]于 2008 年提出用相干移动簇算法(Coherent moving cluster, CMC) 算法来挖掘护航模式, 由于 CMC 算法要为缺失的点生成虚拟点然后对每个时间戳上的所有移动对象聚类, 因此需要较大的时间开销。为了避免上述问题, 采用轨迹简化技术的伴随模式发现算法(Convoy discovery using trajectory simplification, CUTS) 和 CUTS+ 算法分别引入了轨迹简化算法(Douglas-Peucker, DP) 和 DP *^[33] 两个轨迹简化技术简化轨迹, 在得到简化轨迹后对轨迹段进行聚类。Jeung 在 CUTS 的基础上考虑到轨迹简化和距离计算的时

间特性, 提出了 CUTS * 算法来进一步提高发掘效率。

Jeung 等人^[34]于 2008 年提出了简化轨迹上的队伍发现算法(Convoy over simplified trajectory, CoST) 和 CoST * 算法, 通过 DP 算法简化轨迹, 然后对轨迹段聚类, 在 CoST 算法的基础上, 考虑轨迹化简和距离计算的时间特性, 提出 CoST * 算法提高挖掘效率。

Yoon 等人^[35]于 2009 年提出了有效队伍发现算法(Valid convoy discovery algorithm, VCoDA) 算法用于挖掘护航模式。一共包含两个阶段, 第一阶段用 CMC 算法找到候选护航模式, 第二阶段验证每个候选护航模式是否是有效的护航模式。

Aung 等人^[20]于 2012 年提出简单分层算法(Simple slice-by-slice, s³) 挖掘发展护航模式, 这个算法和之前提出的相干移动簇算法类似, 首先对缺失的位置信息进行线性插值, 然后基于密度聚类, 最后再进行查询, 和相干移动簇算法一样, 该算法也要进行大量的 DBSCAN 聚类, 时间开销非常大。因此 Aung 提出了 ID-1 算法, 在对轨迹进行简化后, 利用 TRAJ-DBSCAN 代替原来的遍历扫描, 降低了基于密度聚类的时间开销; 在 ID-1 算法的基础上, Aung 等人又提出了 ID-2 算法, 通过增加剪枝操作进一步优化算法。

Li 等人^[21]于 2010 年提出对象增长(Object-Growth) 算法用于挖掘蜂群模式, 通过将簇空间对象和时间组织成一颗前序树, 并利用剪枝策略加速搜索过程。Apriori 剪枝用于停止不满足时间阈值的子树搜索, 后向剪枝已有对象集合的子集。并在搜索过程中通过前向封闭检测(Forward closure checking) 算法确定是否是封闭蜂群模式。

Yu 等人^[39]于 2010 年提出了随时间增长的簇重组算法(Time growth cluster recombinant algorithm, TGCR) 算法用于挖掘蜂群模式, 该算法首先根据每个时间戳的聚类结果找到相同簇中的所有对象集合, 然后检查每个对象集合及其对应的最大时间集合是否满足模式的定义要求。

Tang 等人^[22]于 2012 年提出了智能封闭算法(Smart-and-closed algorithm) 用于挖掘伴游模式。提出并建立了一种新的移动小团体(Traveling Buddy) 数据结构, 基于该数据结构, 算法不需要考虑每个对象, 能够加速挖掘过程。算法将小团体(Buddy) 对象聚类成簇, 通过建立小团体索引确定相邻时刻簇的共同对象, 挖掘伴游模式。

Li 等人^[23]于 2015 年提出了排组模式挖掘

(PlatoonMiner)算法,用于挖掘群组模式,该算法分为两个阶段,包括预处理阶段和挖掘阶段。预处理阶段,在每个时间戳将所有对象聚类成簇。挖掘阶段,建立搜索树,以深度优先搜索挖掘该搜索树得到候选并存储于前缀表中,并利用四种剪枝规则加速挖掘过程,最后根据条件移除不合格的候选。

(3) 基于区域查询的模式挖掘算法

2006年, Benkert 等人^[17]提出用近似算法来挖掘成群模式。首先将移动对象轨迹映射到高维空间,然后对映射点集建立四叉树(Skip-quadtree)索引^[36],分别采用盒状(Box),管道(Pipe),采样点(Sample-points)三种近似算法进行区域查询。同年, Gudmundsson^[37]针对之前提出的3种没有考虑算法优化问题,且算法时间复杂度是指数级的近似查询算法进行了改进,在改进过程中对成群模式的定义进行了修正,不再要求成群模式的时间是整数值。在采用近似算法时忽略移动对象彼此之间的位置关系而只考虑它们在群体中的时间,将问题转化为找出一个最长的时间区间,要求该时间区间至少包含指定阈值规模的移动对象,并在查找过程中建立扩充二叉树索引。

2007年, Al-Naymat 等人指出之前的成群模式挖掘算法运行时间与模式时长成指数关系,不能很好地解决挖掘长时间序列的成群模式。于是他们对 Gudmundsson 提出的挖掘算法进行改进^[40],将所有轨迹映射到 2K 维空间,得到映射点集 P' 之后再对点集 P' 进行随机投影操作^[41-42],获得在 $\frac{4+2\beta}{\epsilon^2/2-\epsilon^2/3} \log n$ 维空间的投射点集 P'' ,并在 P'' 上进行区域查询。

3.3 聚集运动模式挖掘

聚集运动模式挖掘算法主要分为两种,基于聚类的模式挖掘算法及基于区域查询的模式挖掘算法。表7所示为群体聚集运动模式的挖掘算法、提出时间、文献以及类型的描述。

(1) 基于聚类的模式挖掘算法

Zheng 等人^[24]于2013年提出了检测和划分(Test and divide, TAD)算法,用于挖掘聚合模式,主要包括3个阶段:第一阶段是对数据库中每个时间戳的移动对象聚类得到簇;第二阶段是挖掘空间位置稳定的簇序列,并通过网格索引加速挖掘过程;第三阶段检测和划分算法挖掘满足稳定参与者数量阈值的簇序列。2014年,他们用检测和划分算法,在线挖掘聚合模式,通过移动小团体数据结

构实现快速聚类,根据移动小团体剪枝挖掘群体对象(Crowd),最后通过检测和划分算法挖掘满足稳定参与者数量阈值的簇序列^[43]。

表7 聚集运动模式挖掘算法

Tab. 7 Aggregation moving pattern mining algorithm			
类型	运动模式	提出时间/年	文献
基于聚类的模式挖掘算法	聚合模式	2013	[23]
		2014	[42]
	新增汇聚模式	2014	[43]
		2012	[24]
滚雪球模式	2014	[25]	
基于区域查询的模式挖掘算法	黑洞模式	2015	[28]

Hai 等人^[24]于2012年提出簇增长(ClusterGrowth)算法用于挖掘新增汇聚模式,将搜索空间中所有簇组织为前序树,树节点是空间轨迹对象,并以数字标记深度优先搜索的顺序。同时,提出了两种剪枝规则更近一步缩小搜索空间,最后通过实际最大值(Actual maximum)方法检测是否是最大的新增汇聚模式。

2014年, Guo 等人^[26]提出了聚类扫描(Clustering-and-scanning)算法用于挖掘滚雪球模式,首先对每个时间戳上的所有移动对象聚类,然后扫描之后时间片的每个簇判断是否可扩展,为了提高效率,引入剪枝规则。

Zhang 等人^[44]于2014年提出聚集检索(Gathering retrieving, GR)算法,用于挖掘聚合模式,通过 DBSCAN 聚类得到移动对象簇,并通过移动对象簇构建时空图(Spatio-temporal graph, STG),最后根据最大团理论从时空图上有效且高效挖掘聚合模式。

(2) 基于区域查询的模式挖掘算法

Hong 等人^[29]于2015年提出了两步黑洞检测算法(Two-step black hole detection algorithm),包括两个部分:选择算法和空间扩展算法,用于挖掘黑洞模式和火山模式。根据城市区域建立时空图,并将时空图划分为网格区域,建立时空图索引,选择算法根据网格区域内子图中人群进流量和人群出流量找到可能包含黑洞的子图,再通过空间扩展算法扩展时空子图为黑洞模式。

4 结束语

群体运动模式在时空轨迹数据挖掘研究和轨迹模式研究中具有重要的理论意义和实际应用价值,目前存在着较多的困难和挑战。本文以群体

运动模式的分类和发展为主线,对群体运动模式的类别和发展进行了阐述,介绍了各种群体运动模式之间的区别与联系,对各种现有的群体运动模式挖掘算法进行了综述。

参考文献:

- [1] Jeung H, Yiu M L, Jensen C S. Trajectory pattern mining[M]. New York: Springer, 2011: 143-177.
- [2] Liu Y, Zhao Y, Chen L, et al. Mining frequent trajectory patterns for activity monitoring using radio frequency tag arrays[J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(11): 2138-2149.
- [3] Gidofalvi G, Pedersen T B. Cab-sharing: An effective, door-to-door, on-demand transportation service [C]// 6th European Congress on Intelligent Transport Systems and Services. London: IET, 2007: 8-15
- [4] Sumpter N, Bulpitt A. Learning spatio-temporal patterns for predicting object behaviour[J]. Image and Vision Computing, 2000, 18(9): 697-704.
- [5] Iwase S, Saito H. Tracking soccer player using multiple views[C] // Proceedings of IAPR workshop on Machine Vision Applications. New York: American Society of Mechanical Engineers, 2002: 102-105.
- [6] Imfeld S. Time, points and space—Towards a better analysis of wildlife data in GIS[D]. Switzerland: Geographisches Institut der Universität, 2000.
- [7] Frank A U. Socio-economic units: Their life and motion[J]. Life and Motion of Socio-Economic Units: GISDATA, 2001, 8: 21-34.
- [8] Laube P, Imfeld S. Analyzing relative motion within groups of trackable moving point objects[C]// Second International Conference on Geographic Information Science. New York: ACM, 2002: 132-144.
- [9] Laube P, van Kreveld M, Imfeld S. Finding REMO-detecting relative motion patterns in geospatial life-lines[C]// 11th International Symposium on Spatial Data Handling. Berlin: Springer, 2005: 201-215.
- [10] Wang Y, Lim E P, Hwang S Y. On mining group patterns of mobile users [C]// 14th International Conference on Database and Expert Systems Applications. Berlin: Springer, 2003: 287-296.
- [11] Feuerhake U, Sester M. Mining group movement patterns [C]// Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2013: 520-523.
- [12] Wang S, Wu L, Zhou F, et al. Group pattern mining on moving objects' uncertain trajectories[J]. International Journal of Computers Communications & Control, 2015, 10(3): 428-440.
- [13] Hadjieleftheriou M, Kollios G, Gunopulos D, et al. On-line discovery of dense areas in spatio-temporal databases[M]. Berlin: Springer, 2003:306-324.
- [14] Gaffney S, Smyth P. Trajectory clustering with mixtures of regression models[C]// Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. California: ACM, 1999: 63-72.
- [15] Ester M, Kriegel H P, Sander J, et al. Incremental clustering for mining in a data warehousing environment[C]// Proceedings of 24th International Conference on Very Large Data Bases. New York: ACM, 1998, 323-333.
- [16] Kalnis P, Mamoulis N, Bakiras S. On discovering moving clusters in spatio-temporal data[J]. Lecture Notes in Computer Science, 2005, 3633:923.
- [17] Benkert M, Gudmundsson J, Hübner F, et al. Reporting flock patterns[J]. Computational Geometry, 2010, 41(3):111-125.
- [18] Gudmundsson J, van Kreveld M, Speckmann B. Efficient detection of motion patterns in spatio-temporal data sets[C]// Proceedings of the 12th annual ACM international workshop on Geographic information systems. New York: ACM, 2004: 250-257.
- [19] Jeung H, Yiu M L, Zhou X, et al. Discovery of convoys in trajectory databases[J]. Proceedings of the Very Large Database Endowment, 2008, 1(1): 1068-1080.
- [20] Aung H H, Tan K L. Discovery of evolving convoys [C]// International Conference on Scientific and Statistical Database Management. Heidelberg: Springer, 2010: 196-213.
- [21] Li Z, Ding B, Han J, et al. Swarm: Mining relaxed temporal moving object clusters[J]. Proceedings of the Very Large Database Endowment, 2010, 3(1): 723-734.
- [22] Tang L A, Zheng Y, Yuan J, et al. On discovery of traveling companions from streaming trajectories [C]// Proceedings of the 28th International Conference on Data Engineering. New York: IEEE, 2012: 186-197.
- [23] Li Y, Bailey J, Kulik L. Efficient mining of platoon patterns in trajectory databases[J]. Data & Knowledge Engineering, 2015, 100: 167-187.
- [24] Hai P N, Ienco D, Poncelet P, et al. Mining time relaxed gradual moving object clusters[C]// Proceed-

- ings of the 20th International Conference on Advances in Geographic Information Systems. California: ACM, 2012; 478-481.
- [25] Zheng K, Zheng Y, Yuan N J, et al. On discovery of Gathering patterns from trajectories[C]// IEEE 29th International Conference on Data Engineering. New York: IEEE, 2013; 242-253.
- [26] Guo L, Huang G, Ding Z. Efficient detection of emergency event from moving object data streams [C]// International Conference on Database Systems for Advanced Applications. Berlin: Springer, 2014; 422-437.
- [27] Li Z, Xiong H, Liu Y, et al. Detecting black hole and volcano patterns in directed networks[C]// Proceedings of the 2010 IEEE International Conference on Data Mining. New York: IEEE, 2010; 294-303.
- [28] Li Z, Xiong H, Liu Y. Mining black hole and volcano patterns in directed graphs: A general approach [J]. *Data Mining and Knowledge Discovery*, 2012, 25(3): 577-602.
- [29] Hong L, Zheng Y, Yung D, et al. Detecting urban black holes based on human mobility data[C]// Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2015; 35-44.
- [30] Solera F, Calderara S, Cucchiara R. Learning to identify leaders in crowd [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2015; 43-48.
- [31] Andersson M, Gudmundsson J, Laube P, et al. Reporting leadership patterns among trajectories[C]// Proceedings of the 2007 ACM Symposium on Applied computing. New York: ACM, 2007; 3-7.
- [32] Hwang S Y, Liu Y H, Chiu J K, et al. Mining mobile group patterns: A trajectory-based approach [C]// Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Heidelberg: Springer, 2005; 713-718.
- [33] Meratnia N, Rolf A. Spatiotemporal compression techniques for moving point objects[C]// Proceeding of International Conference on Extending Database Technology. Berlin: Springer, 2004; 765-782.
- [34] Jeung H, Shen H T, Zhou X. Convoy queries in spatio-temporal databases [C]// Proceeding of 2008 IEEE 24th International Conference on Data Engineering. New York: IEEE, 2008; 1457-1459.
- [35] Yoon H, Shahabi C. Accurate Discovery of Valid Convoys from Moving Object Trajectories[C]// Proceeding of IEEE International Conference on Data Mining. New York: IEEE, 2009; 636-643.
- [36] Eppstein D, Goodrich M T, Sun J Z. The skip quadtree: A simple dynamic data structure for multi-dimensional data [C]// Proceedings of the twenty-first annual symposium on Computational geometry. New York: ACM, 2005; 296-305.
- [37] Gudmundsson J, van Kreveld M. Computing longest duration flocks in trajectory data[C]// Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems. New York: ACM, 2006; 35-42.
- [38] Wang Y, Lim E P, Hwang S Y. Efficient algorithms for mining maximal valid groups [J]. *The VLDB Journal—The International Journal on Very Large Data Bases*, 2008, 17(3): 515-535.
- [39] Yu Y, Wang Q, Kuang J, et al. TGCR: An efficient algorithm for mining swarm in trajectory databases [C]// Proceeding of 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services. New York: IEEE, 2011; 90-95.
- [40] Al-Naymat G, Chawla S, Gudmundsson J. Dimensionality reduction for long duration and complex spatio-temporal queries [C]// Proceedings of the 2007 ACM Symposium on Applied Computing. New York: ACM, 2007; 393-397.
- [41] Bagnall A, Keogh E, Lonardi S, et al. A bit level representation for time series data mining with shape based similarity [J]. *Data Mining and Knowledge Discovery*, 2006, 13(1): 11-40.
- [42] Bingham E, Mannila H. Random projection in dimensionality reduction: applications to image and text data [C]// Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2001; 245-250.
- [43] Zheng K, Zheng Y, Yuan N J, et al. Online discovery of gathering patterns over trajectories [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(8): 1974-1988.
- [44] Zhang J, Li J, Wang S, et al. On retrieving moving objects gathering patterns from trajectory data via spatio-temporal graph [C]// 2014 IEEE International Congress on Big Data. New York: IEEE, 2014; 390-397.